

IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V–J and V–D–J rearrangement analysis

Véronique Giudicelli¹, Denys Chaume¹ and Marie-Paule Lefranc^{1,2,*}

¹IMGT, the international ImMunoGeneTics information system[®], Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Institut de Génétique Humaine IGH, UPR CNRS 1142, 141 rue de la Cardonille, F-34396 Montpellier Cedex 5, France and ²Institut Universitaire de France

Received February 12, 2004; Revised and Accepted April 1, 2004

ABSTRACT

IMGT/V-QUEST, for 'V-QUery and STandardization', is an integrated software program which analyses the immunoglobulin (IG) and T cell receptor (TR) rearranged nucleotide sequences. The extraordinary diversity of the IG and TR repertoires (10^{12} antibodies and 10^{12} TR per individual) results from several mechanisms at the DNA level: the combinatorial diversity of the variable (V), diversity (D) and joining (J) genes, the N-diversity and, for IG, the somatic mutations. IMGT/V-QUEST identifies the V, D and J genes and alleles by alignment with the germline IG and TR gene and allele sequences of the IMGT reference directory. IMGT/V-QUEST delimits the structurally important features, frameworks and complementarity-determining regions (the last of these forming the antigen binding site), on the basis of the IMGT unique numbering. The tool localizes the somatic mutations of the IG rearranged sequences. IMGT/V-QUEST also dynamically displays a graphical two-dimensional representation, or IMGT Collier de Perles, of the IG and TR variable regions. Moreover, IMGT/V-QUEST can interact with IMGT/JunctionAnalysis for the detailed description of the V–J and V–D–J junctions, and with IMGT/PhyloGene for the construction of phylogenetic trees. IMGT/V-QUEST is currently available for human and mouse, and partly for non-human primates, sheep, chondrichthyes and teleostei. IMGT/V-QUEST is freely available at <http://imgt.cines.fr>.

INTRODUCTION

IMGT, the international ImMunoGeneTics information system[®] (<http://imgt.cines.fr>) (1,2), created in 1989 by the

Laboratoire d'ImmunoGénétique Moléculaire (LIGM) (Université Montpellier II and CNRS) at Montpellier, France, is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC), immunoglobulin superfamily and related proteins of the immune system (RPI) of human and other vertebrate species. IMGT comprises six databases (3–5), 10 interactive tools (6–9) and 8000 HTML pages of genomic, genetic and structural data related to IG, TR, MHC and RPI (2,5). In particular, IMGT deals with all the IG and TR nucleotide sequences published in the generalist databases DDBJ, EMBL and GenBank (10–12). Most of these are rearranged sequences resulting from the very complex synthesis that includes combinatorial diversity rearrangements between the variable (V), diversity (D) and joining (J) genes, the N-diversity mechanism and, for IG, the somatic mutations [for review see (13,14)]. Indeed, these three mechanisms create the huge diversity of the IG and TR repertoires (10^{12} antibodies and 10^{12} TR per individual). One major challenge for IMGT is to provide users with a detailed and accurate characterization of these sequences according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY concepts (15). This objective is achieved with IMGT/V-QUEST (V-QUery and STandardization), an integrated on-line software program which analyses IG and TR rearranged nucleotide sequences. IMGT/V-QUEST identifies the V, D and J genes and alleles in the rearranged V–J and V–D–J sequences. This step is performed by aligning and comparing the input sequence with the reference sequence sets of the germline (not rearranged) IG and TR genes and alleles, defined according to the IMGT criteria and grouped in the IMGT reference directory sets. IMGT/V-QUEST numbers the translated amino acids of the IG and TR variable region on the basis of the IMGT unique numbering (16) and delimits the region's structurally important features,

*To whom correspondence should be addressed. Tel: +33 4 99 61 99 65; Fax: +33 4 99 61 99 01; Email: lefranc@ligm.igh.cnrs.fr

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated.

© 2004, the authors

the three frameworks according to the IMGT unique numbering (FR-IMGT) (17) and the three complementarity-determining regions according to the IMGT unique numbering (CDR-IMGT) (the last of these forming the antigen binding site). IMGT/V-QUEST dynamically displays a graphical two-dimensional (2D) representation, or IMGT Collier de Perles (16,18), of the variable region, clearly delimiting the structural CDR-IMGT loops and FR-IMGT beta sheet strands. Moreover, IMGT/V-QUEST provides users with optional and complementary analyses, such as detailed description of the V–J or V–D–J JUNCTION between the V 3' end and the J 5' end using IMGT/JunctionAnalysis, and the construction of phylogenetic trees computed with IMGT/PhyloGene (7).

IMGT/V-QUEST is dedicated to the analysis of V–J and V–D–J sequences of any vertebrate species provided that their germline IG and TR genes are known, have been expertly analysed by IMGT and are in IMGT Repertoire (<http://imgt.cines.fr>). IMGT/V-QUEST is currently available for human and mouse, and partly for non-human primates, sheep, chondrichthyes and teleostei. IMGT/V-QUEST is freely available from the IMGT home page (<http://imgt.cines.fr>).

IMGT/V-QUEST INTEGRATED ON-LINE TOOL

Algorithm, reference directory and ontology

An input sequence is entered by the user into the IMGT/V-QUEST tool as an IG or a TR sequence. IMGT/V-QUEST first identifies the locus to which the sequence belongs by comparison with a pool of test V-REGION sequences (from the IGH, IGK and IGL loci for an IG sequence, and from the TRA, TRB, TRG and TRD loci for a TR sequence) (13,14).

IMGT/V-QUEST then delimits the V-REGION in the input sequence and numbers the codons by alignment, for a given locus, with the master sequence set. The master sequence set comprises one representative of each V-REGION, with different FR-IMGT and CDR-IMGT lengths. Gaps are introduced into the input sequence at positions which are unoccupied in comparison with the master sequences.

In a next step, IMGT/V-QUEST identifies the V, D and J genes and alleles by alignment of the input sequence with the IMGT reference directory. The IMGT/V-QUEST reference directory sets are constituted, for a given species and for a given locus, by sets of sequences which contain the V-REGION, D-REGION and J-REGION alleles isolated from the functional and open reading frame (ORF) allele IMGT reference sequences for the V, D and J genes, respectively. By definition, these sets contain one sequence for each allele. The names of these allele sequences are shown in red in 'Alignments of alleles' in IMGT Repertoire (<http://imgt.cines.fr>). Exceptionally, the IMGT/V-QUEST reference directory sets may include sequences isolated from pseudogene allele IMGT reference sequences [indicated with (P) following the allele name]. The set of sequences from the IMGT reference directory used for IMGT/V-QUEST can be downloaded in FASTA format from the IMGT/V-QUEST page. The IMGT/V-QUEST program first aligns the V-REGION, then the J-REGION (on the sequence located downstream of the 3' deduced end of the V-REGION) and finally the D-REGION (on the sequence between the 3' deduced end of the

V-REGION and the 5' deduced end of the J-REGION). IMGT/V-QUEST delimits the three FR-IMGT and the three CDR-IMGT of the V-REGION, and the JUNCTION, based on the DESCRIPTION and NUMEROTATION concepts of IMGT-ONTOLOGY (15). The JUNCTION encompasses the sequence between the V-REGION conserved cysteine (2nd-CYS) at position 104 and the J-REGION conserved phenylalanine (J-PHE) (for the IG light chains and the TR chains) or tryptophan (J-TRP) (for the IG heavy chains) at position 118. The analysis of the JUNCTION by IMGT/JunctionAnalysis has been integrated into IMGT/V-QUEST. This option is currently selected by default because it has the advantage of providing a detailed description of the N-diversity regions [N-REGION(s)].

An IMGT/V-QUEST query

An IMGT/V-QUEST query consists of two steps:

- (i) The user selects the antigen receptor (Immunoglobulin or T cell receptor) and the species on the IMGT/V-QUEST Search page.
- (ii) On the next page, the user has the options not to include IMGT/JunctionAnalysis and to give a name to the input sequence. The user provides the input nucleotide sequence by one of the three following means: typing the sequence, copying and pasting it, or uploading it by giving the path to a local file containing the sequence in FASTA format.

IMGT/V-QUEST will give a message indicating an error or aberrant results for pseudogene sequences with nucleotide insertions or deletions (to which the IMGT unique numbering cannot be applied), for too short partial sequences (for which the V-GENE cannot be identified), for reverse sequences, and for sequences containing a cluster of V-GENEs, or sequences with too long 5'- or 3'-untranslated region.

IMGT/V-QUEST output

The output of IMGT/V-QUEST shows six major results:

- (i) Alignments for V-GENE, D-GENE and J-GENE. IMGT/V-QUEST displays the alignments of the input sequence with the closest five V and D genes and alleles, and the closest three J genes and alleles, of the IMGT reference directory (Figure 1). The best alignment score indicates the gene and allele most likely to be involved in the rearrangement. Figure 1A shows that for the input sequence, the V–D–J rearrangement involves the variable IGHV3-30 * 04 allele, the diversity IGHD4-23 * 01 allele, and the joining IGHJ4 * 02 allele.
- (ii) Translation of the JUNCTION. IMGT/V-QUEST displays the translation of the V–J, or V–D–J, JUNCTION, delimited by the V-REGION 2nd-CYS at position 104 and the J-REGION J-PHE or J-TRP at position 118 (both positions according to the IMGT unique numbering) (16). The display also includes the three amino acids at positions 119, 120 and 121, which are part of the characteristic J-REGION motif (F/W–G–X–G). In the default option, the alignment with the D gene is replaced by the analysis of the JUNCTION by IMGT/JunctionAnalysis. In this case, the alignment with the best D gene and allele is

A

THANK YOU
for using **IMGT/V-QUEST**

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

ImMunoGeneTics
Information system®

<http://imgt.cines.fr>

Sequence compared with the **Human Ig set** from the **IMGT reference directory**

Hyphens (-) show nucleotide identity, dots (.) represent gaps

Alignment for V-GENE

input	score	sequence
I06615 IGHV3-30*04	1200	CAGGTGCAAGCTGGTGGAGCTCTGGGGGA...GGCGTGGTCCAGCCTGGGAGGTCCTGAGACTCTCTGTGCAAGCCTCTGGATTCCGCTTCGAK
X92283 IGHV3-30*3*01	1191	C-----A-----AG
M77324 IGHV3-30*14	1191	C-----A-----AG
M77300 IGHV3-30*09	1191	C-----A-----AG
M83134 IGHV3-30*01	1191	C-----A-----AG

Alignment for D-GENE

input	score	sequence
X97051 IGHD4-23*01	32	---A--T--
X97051 IGHD4-17*01	23	C---TA-
J00256 IGHD7-27*01	21	--G--
X97051 IGHD3-16*02	14	-C--T-TA--
X93614 IGHD3-16*01	14	--C-T-TA--

Alignment for J-GENE

input	score	sequence
X86355 IGHJ4*02	199-C-----G
J00256 IGHJ4*01	190-AA--A-----C-----G
M25625 IGHJ4*03	181GC-----AA--G-----C-----G

Translation of the JUNCTION

input	104	118
	C A R D G D P I G Y Y F D Y W G P G	
	TGT GCG AGA GAT GGT GAC CCC ATA GGA TAT TAC TTT GAC TAC TGG GCG CCG GCG	

B

This release of IMGT/V-QUEST uses **IMGT/JunctionAnalysis** for the analysis of the JUNCTION

Sequence compared with the **Human Ig set** from the **IMGT reference directory**

Hyphens (-) show nucleotide identity, dots (.) represent gaps

Alignment for V-GENE

input	score	sequence
I06615 IGHV3-30*04	1200	CAGGTGCAAGCTGGTGGAGCTCTGGGGGA...GGCGTGGTCCAGCCTGGGAGGTCCTGAGACTCTCTGTGCAAGCCTCTGGATTCCGCTTCGAK
X92283 IGHV3-30*3*01	1191	C-----A-----AG
M77324 IGHV3-30*14	1191	C-----A-----AG
M77300 IGHV3-30*09	1191	C-----A-----AG
M83134 IGHV3-30*01	1191	C-----A-----AG

Alignment for J-GENE

input	score	sequence
X86355 IGHJ4*02	199-C-----G
J00256 IGHJ4*01	190-AA--A-----C-----G
M25625 IGHJ4*03	181GC-----AA--G-----C-----G

Results of IMGT/JunctionAnalysis

Analysis of the JUNCTION

Input V name	V-REGION	D-REGION	N2	J-REGION	J name	D name	Vmut	Dmut	Jmut	
#1 input	IGHV3-30*04	tgtgagagagatgtgagagag	ataggatgata	tactttgactactgg	IGHJ4*02	IGHD4-23*01	0	2	0

Input Ngs

#1 input 2/9

Translation of the JUNCTION

104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	frame	length
C	A	R	D	G	D	P	I	G	Y	Y	F	D	Y	W		
#1 input	tgt	gcg	aga	gat	ggt	gac	ccc	ata	gga	tat	tac	ttt	gac	tac	tg	13

Figure 1. IMGT/V-QUEST results. (A) Alignments for V-GENE, D-GENE and J-GENE, and translation of the JUNCTION. The input sequence is aligned with the closest V, D and J genes and alleles of the IMGT reference directory set. The IMGT/LIGM-DB accession number (identical to the DDBJ/EMBL/GenBank accession number), the IMGT gene name and allele are indicated for each reference sequence. Dashes indicate nucleotide identity. Dots indicate gaps according to the IMGT numbering or nucleotides that are not taken into account for the alignment (16). The JUNCTION extends from the 2nd-CYS (C) at position 104 to J-TRP (W) at position 118. (B) The default option of IMGT/V-QUEST: the alignment for D-GENE is replaced by the results of IMGT/JunctionAnalysis.

displayed with a detailed description of the N-diversity regions (Figure 1B).

- (iii) Alignment with the FR-IMGT and CDR-IMGT delimitations. IMGT/V-QUEST shows the alignment of the input nucleotide sequence with the closest V-REGIONS and with the FR-IMGT and CDR-IMGT delimitations (FR1-, CDR1-, FR2-, CDR2-, FR3- and germline CDR3-IMGT). This representation allows the user to easily compare the sequences and to locate somatic mutations (Figure 2A).
- (iv) Translation. IMGT/V-QUEST provides the translation of the input sequence. The nucleotide and deduced amino acid sequences are shown with the FR-IMGT and CDR-IMGT delimitations (FR1-, CDR1-, FR2-, CDR2-, FR3- and rearranged CDR3-IMGT) (Figure 2B).
- (v) 'Input' V-REGION. IMGT/V-QUEST provides the input V-REGION nucleotide sequence in FASTA format, with gaps inserted according to the IMGT unique numbering (16). By clicking on the button 'Analyse this sequence with IMGT/PhyloGene', this formatted nucleotide sequence is automatically added to the reference sequence set of IMGT/PhyloGene for phylogenetic studies (7) (Figure 2C).
- (vi) IMGT Collier de Perles. IMGT/V-QUEST automatically displays a graphical 2D representation, or IMGT Collier de Perles, of the input V-REGION amino acid sequence (16). This representation is particularly useful for visualizing the beta sheet strands that compose the FR-IMGT, the loops of the CDR1-IMGT and CDR2-IMGT that are part of the antigen binding site, and the positions of the conserved amino acids (Figure 2D).

Implementation

IMGT/V-QUEST was implemented using Java 2 and cgi-bin programs. It uses, for the sequence alignments, the DNAPLOT program, an alignment tool component of IMGT developed by Hans-Helmar Althaus and Werner Müller (Institut für Genetik, Köln, Germany) (19), and for the IMGT Collier de Perles representations, the IMGT/Collier de Perles tool developed by Gérard Mennessier (Montpellier, France) (20).

DISCUSSION AND CONCLUSION

IMGT/V-QUEST (<http://imgt.cines.fr>) is the first tool available on the Web that identifies and delimits the V, D and J genes and alleles within IG or TR nucleotide rearranged sequences, that delimits the frameworks and the complementarity-determining regions, and that displays a graphical 2D representation of the variable region. Moreover, IMGT/V-QUEST interacts with additional IMGT tools for more detailed analysis, via an option on the IMGT/V-QUEST Search page for a detailed analysis of the JUNCTION using IMGT/Junction-Analysis, and via a link to IMGT/PhyloGene on the result page for the phylogenetic analysis of the variable region of the input sequence.

Although IMGT/V-QUEST is mostly used to characterize rearranged V-J and V-D-J sequences, it can also be used to identify genes in germline V-GENES. To the best of our knowledge, IMGT/V-QUEST is the first Web tool that

displays such detailed and accurate analysis of nucleotide rearranged sequences from antigen receptors. The information provided by IMGT/V-QUEST is of much value to clinicians and biological scientists studying the repertoire of the immunoglobulins and T cell receptors in physiological and pathological immune responses (autoimmune diseases, leukaemias, lymphomas, myelomas, infectious diseases, etc.) (21,22), for evolution studies and for antibody engineering.

AVAILABILITY AND CITATION

Authors who use IMGT/V-QUEST are encouraged to cite this article and to quote the IMGT home page URL, <http://imgt.cines.fr>.

ACKNOWLEDGEMENTS

We thank Gérard Lefranc for helpful discussion. We are deeply grateful to the IMGT team for its expertise and constant motivation and specially to our curators for their hard work and enthusiasm. IMGT is funded by the European Union's 5th PCRDT programme (QLG2-2000-01287), the Centre National de la Recherche Scientifique (CNRS) and the Ministère de la Recherche et de l'Education Nationale.

REFERENCES

- Lefranc, M.-P. (2003) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, **31**, 307–310.
- Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bosc, N., Folch, G., Guiraudou, D., Jabado-Michaloud, J., Magris, S., Scaviner, D., Thouvenin, V., Combres, K., Girod, D., Jeanjean, S., Protat, C., Yousfi Monod, M., Duprat, E., Kaas, Q., Pommie, C., Chaume, D. and Lefranc, G. (2003) IMGT-ONTOLOGY for immunogenetics and immunoinformatics (<http://imgt.cines.fr>). *In Silico Biol.*, **4**, 0004. <http://www.bioinfo.de/isb/2003/04/2004/>
- Lefranc, M.-P. (2003) IMGT® databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, <http://imgt.cines.fr>. *Leukemia*, **17**, 260–266.
- Lefranc, M.-P. (2003) IMGT, the international ImMunoGeneTics information system®, <http://imgt.cines.fr>. In Bock, G. and Goode, J. (eds), *Immunoinformatics: Bioinformatics Strategies for Better Understanding of Immune Function*, Novartis Foundation Symposium 254. John Wiley and Sons, Chichester, pp. 126–136, discussion pp. 136–142, 216–222, 250–252.
- Lefranc, M.-P. (2004) IMGT-ONTOLOGY and IMGT databases, tools and web resources for immunogenetics and immunoinformatics. *Mol. Immunol.*, **40**, 647–659.
- Lefranc, M.-P. (2003) IMGT, the international ImMunoGeneTics information system®, <http://imgt.cines.fr>. In Lo, B.K.C. (ed.), *Antibody Engineering: Methods and Protocols. Methods in Molecular Biology*, 2nd edn. Humana Press, Totowa, NJ, Vol. 248, chap 3, pp. 27–49.
- Elemento, O. and Lefranc, M.-P. (2003) IMGT/PhyloGene: an on-line tool for comparative analysis of immunoglobulin and T cell receptor genes. *Dev. Comp. Immunol.*, **27**, 763–779.
- Kaas, Q., Ruiz, M. and Lefranc, M.-P. (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res.*, **32**, D208–D210.
- Baum, T.P., Pasqual, N., Thuderoz, F., Hierle, V., Chaume, D., Lefranc, M.-P., Jouvin-Marche, E., Marche, P.N. and Demongeot, J. (2004) IMGT/GeneInfo: enhancing V(D)J recombination database accessibility. *Nucleic Acids Res.*, **32**, D51–D54.

10. Miyazaki,S., Sugawara,H., Ikeo,K., Gojobori,T. and Tateno,Y. (2004) DDBJ in the stream of various biological data. *Nucleic Acids Res.*, **32**, D31–D34.
11. Kulikova,T., Aldebert,P., Althorpe,N., Baker,W., Bates,K., Browne,P., van den Broek,A., Cochrane,G., Duggan,K., Eberhardt,R. *et al.* (2004) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **32**, D27–D30.
12. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2004) GenBank: update. *Nucleic Acids Res.*, **32**, D23–D26.
13. Lefranc,M.-P. and Lefranc,G. (2001) *The Immunoglobulin Facts Book*. Academic Press, London, 458 pages.
14. Lefranc,M.-P. and Lefranc,G. (2001) *The T Cell Receptor Facts Book*. Academic Press, London, 398 pages.
15. Giudicelli,V. and Lefranc,M.-P. (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics*, **15**, 1047–1054.
16. Lefranc,M.-P., Pommié,C., Ruiz,M., Giudicelli,V., Foulquier,E., Truong,L., Thouvenin-Contet,V. and Lefranc,G. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, **27**, 55–77.
17. Pommié,C., Levadoux,S., Sabatier,R., Lefranc,G. and Lefranc,M.-P. (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J. Mol. Recogn.*, **17**, 17–32.
18. Ruiz,M. and Lefranc,M.-P. (2002) IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics*, **53**, 857–883.
19. Giudicelli,V., Chaume,D., Mennessier,G., Althaus,H.-H., Müller,W., Bodmer,J., Malik,A. and Lefranc,M.-P. (1998) IMGT, the international ImMunoGeneTics database: a new design for immunogenetics data access. In Cesnik,B., McGray,A.T. and Scherer,J.R. (eds), *Medinfo'1998*. IOS Press, Amsterdam, pp. 351–355.
20. Ruiz,M., Mennessier,G., Giudicelli,V., Chaume,D., Lefranc,G. and Lefranc,M.-P. (2000) Integration of T cell receptor structural data into IMGT, the international ImMunoGeneTics database. *The European Federation for Immunogenetics 14th European Histocompatibility Conference*, Montpellier, France, 4–7 April 2000. *Human Immunol.*, **61**(Suppl. 1), S150, Abstract 286.
21. Zwick,M.B., Kelleher,R., Jensen,R., Labrijn,A.F., Wang,M., Quinnan,G.V., Jr, Parren,P.W., and Burton,D.R. (2003) A novel human antibody against human immunodeficiency virus type 1 gp120 is V1, V2, and V3 loop dependent and helps delimit the epitope of the broadly neutralizing antibody immunoglobulin G1 b12. *J. Virol.*, **77**, 6965–6978.
22. Chardes,T., Chapal,N., Bresson,D., Bes,C., Giudicelli,V., Lefranc,M.-P. and Peraldi-Roux,S. (2002) The human anti-thyroid peroxidase autoantibody repertoire in Graves' and Hashimoto's autoimmune thyroid diseases. *Immunogenetics*, **54**, 141–157.